An Assessment of Emission Factor and Inventory Uncertainty

Ronald E. Myers
Emissions Monitoring and Analysis Division (MD-14)
U.S. Environmental Protection Agency
Research Triangle Park, North Carolina 27711

ABSTRACT

Emission factors that are used to develop emission inventories are typically single values that are arithmetic averages of all source test data that can be obtained. Many of these emission factors are contained in AP-42.[1] Most of the emission factors presented in AP-42 have a quality rating from "A" to "E" suggesting the quality associated with them. The "A" rating being the best quality and the "E" rating being the poorest. This quality rating represents the developers' estimate of the overall assessment of how good a factor is, based on both the quality of the test(s) or information that is the source of the factor and on how well the factor represents the emission source. This paper describes two standard statistical methods that can be used to assess the uncertainty associated with an emission factor. The application of these two methods to assess three likely populations with characteristics similar to many source categories that have been recently evaluated. In addition, this paper presents a method for using these estimates of uncertainty to assess quantitatively the uncertainty in applying this emission factor to an emission inventory. Some potential uses of this uncertainty information are suggested.

INTRODUCTION

Since 1972 Emission Factors presented in AP-42 have included a letter that represents the quality of that value. Guidelines exist which establish some level of consistency in assigning these quality ratings.[2] However, even with these guidelines, the criteria for assigning these letter grades are somewhat subjective and dependent on the engineering judgement, confidence and interpretations of the person developing the factor. Although many people assume that these ratings provide information about the accuracy of the factors presented, these ratings only provide gross indicators of potential biases and lack of supporting data. Variability inherent in the emission factors is not usually included in the assignment of the factor rating. As a result, emission factors with "A" ratings may have more uncertainty than another factor that has a "C" rating. Most often, the users of these emission factors do not care about the quality ratings. This may be appropriate since no reasonable option may be available but to use the factor despite the uncertainties and biases in the factor. Some users assume a level of uncertainty for factors. Typically the assumptions range from a 10 percent accuracy for an "A" rated factor to an order of magnitude estimate for an E rated factor. A growing desire by many is that the uncertainty of emission factors and the inventories built from these emission factors be assessed.[3] The need for this assessment comes from the need to evaluate the quality of an overall inventory, the uncertainty in a modeled event, and the ranking of work to improve an emission inventory. Although still subjective, the Data Attribute Rating System (DARS) provides some insight into some of these needs.[4]

METHODS AVAILABLE TO ASSESS UNCERTAINTY

Most statistical analyses are based upon the premise that the population analyzed is of normal or near normal distribution. Normally distributed populations are symmetrical around their mean and can be adequately described with their mean and their standard deviation. Knowing these two values, the distribution of values within the population can be described using the standard normal distribution. In a normally distributed population, 68% of the population is within one standard deviation of the mean.

Additionally, 95% of the population is within 1.4 standard deviations and 99% of the population is within 2.6 standard deviations of the mean. The number used to multiply by the standard deviations are called the Z values and can be obtained from a cumulative normal distribution table.

Except in rare instances, one cannot measure every member of a population. Therefore, there is some uncertainty that the mean of the samples from the population are the same as population mean. In addition the population standard deviation is not usually known. The Students t distribution provides a method to estimate the Confidence limits for the population mean when the population standard deviation is not known. Besides probability percentages and multipliers (t values), the Students t includes the number of degrees of freedom.[5] When n samples are obtained from a population, the degrees of freedom are n-1 and the confidence limits for μ are calculated using the following pair of inequalities:

$$\overline{X} - t\,s\,/\sqrt{n} \le \mu \le \overline{X} + t\,s\,/\sqrt{n} \qquad \text{(Equation 1)}$$

where:

$\overline{X}$	=	the estimate of the population mean
t	=	the t value associated with the probability from the Students t distribution for n-1 degrees of freedom
s	=	the sample standard deviation of the n values
μ	=	the population mean

The values calculated are called the confidence limits for $\overline{X}$. As can be seen, the confidence is inversely proportional to the square root of the number of samples. The t values for a given percentage decrease with increasing degrees of freedom until at an infinite degree of freedom, the t values are equal to the Z values.

As stated above, these methods to estimate uncertainty are based upon samples from a population that is normally or near normally distributed. Several graphical and statistical methods are available to evaluate whether a population is normally distributed. Most involve the comparison of the frequency distribution of the population with the expected frequency distribution for a Normal population. Other, simpler clues are available to identify a population that is not normally distributed. Besides being symmetrical about the mean, the median and mode of normally distributed populations are the same as the mean. The mean and median are standard descriptive statistics that help characterize a population. The standard deviation and skewness are two other descriptive statistics that can provide information on whether the population is symmetrical. The standard deviation provides information on the spread of the data. Skewness is an indicator of the asymmetry of the population. Symmetrical populations have skewness near zero. Higher values of skewness show the magnitude of the asymmetry. Many emission factors where there are ten or more supporting data have positive skewness with values from 2 to more than 30. In addition, many emission factors have sample standard deviations that approach the value of the mean.[6, 7, 8, 9, 10, 11] Also there are some factors that have sample standard deviations that exceed the mean by factors of two or more. If a population were symmetrical with a sample standard deviation comparable to the mean, then about 34% of the population would be less than zero. It is unlikely that any source that we would develop emission factors for would have any facilities with negative values. Although the more complicated methods that are available can be used to decide whether a population is non normal in distribution, it is believed that this is sufficient evidence that the populations from which these emission factors were determined are not normally distributed.

Sometimes transformations are used to convert non normal populations to a more normal distribution for statistical analysis. The difficulties with these transformations are the calculational complexities in returning to the original space where the original measurements were made. One example that shows this involves sampling from a log normally distributed population. For positively skewed

populations, the arithmetic mean of the population (and samples from the population) is higher than log transforming the population to achieve normal distribution, obtaining the mean and then taking the anti logarithm. Simulations are sometimes used to analyze populations of non normal distributions to avoid these calculational complexities. Different types of simulations include Monte-Carlo simulations and bootstrapping techniques. Bootstrapping techniques involve the repetitive sampling of an existing set of data to infer conclusions about what the data show. Mont-Carlo simulation is the development of a population with distribution like the population from which samples are drawn and then repeatedly sampling from that population to infer conclusions about the actual data obtained from the population.

METHODOLOGY

To assess the uncertainty of non-normally distributed emission factors, Mont-Carlo simulation was used to estimate the confidence intervals for the mean of from one to forty samples. Three log normal distributions were developed to simulate the range of the non normal distributions that appear to exist in the source categories that have been investigated. Although the types of emission factor population distributions that may exist are not known, the log normal distribution was selected as a beginning. It was selected since generating it in spreadsheet programs is easy, it is positively skewed, has no values below zero and it easy for many engineers and environmental scientists to understand. For simplicity of calculations and transferability to other situations, each of the simulated populations had an arithmetic mean of 1.0. One population with a standard deviation of 0.5 was developed to represent a population that might be assumed to be normally distributed since only 2% of the population would fall below zero. One population was developed with a standard deviation of 1.0 to represent a common variability seen in the development of emission factors where there are more than ten sets of data. One population was developed with a standard deviation of 1.8 to represent a higher, although not uncommon, variability than has been seen so far.

The method used to develop the simulated populations use the following two relationships between a log normally distributed population and the transformed population.[12]

$$E[X] = e^{\mu + \frac{\sigma^2}{2}} \quad \text{and,} \qquad \text{(Equation 3)}$$
$$VAR[X] = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1) \qquad \text{(Equation 4)}$$

where:

$E[X]$	=	the arithmetic mean of the population to be developed
$Var[X]$	=	the variance (standard deviation squared) of the population to be developed
μ	=	the mean of the log transformed population
σ	=	the standard deviation of the log transformed population

These equations were solved to find the μ's and σ's to generate normally distributed populations. When the individual values of these normally distributed populations were exponentiated, the resulting populations would have means of 1.0 and standard deviations of 0.5, 1.0 and 1.8. Table 1 presents some descriptive statistics of the three populations in both normal and log space. Table 2 presents the cumulative frequency distributions of these log-normal populations.

Each population was randomly ordered and forty sequential samples were drawn. This was repeated 5,100 times. For each repetition, 40 means and 39 sample standard deviations were determined. The 5,100 sets of 40 means were used to find the average estimate of the mean and the cumulative distributions of the means. The cumulative distributions were used to find the confidence intervals associated with estimating the mean of the parent population using from one to forty random samples.

The 5,100 sets of 39 sample standard deviations were used to find the average sample standard deviation using from two to forty random samples. Tables 4 through 9 present some descriptive statistics for these sample means and sample standard deviations. Tables 10 through 15 present the cumulative distributions and the confidence intervals for the means.

RESULTS

Tables 4, 6 and 8, show that the sample means were unbiased estimators of the population mean and on average is neither too high nor too low. Following statistical theory, the standard deviation of the sample means decrease with increasing sample size and is equal to the population standard deviation divided by the square root of the sample size. However, the result of the random sampling that are presented in Tables 5, 7 and 9 show that the sample standard deviations underestimate the population standard deviation. Some underestimation is expected even for a normally distributed population. This underestimation can be estimated by using the Chi-square values for 50% probability of a greater value. The underestimation predicted for a normally distributed population using the Chi Square values and for the three log-normally distributed populations are presented in Table 3. The underestimation for the log-normal distributions is generally greater than for a normal distribution. Although not presented, it should be noted that log transforming the individual samples, determining the means and standard deviations in log space and using Equations 3 and 4 to estimate the population standard deviation in normal space does not produce this underestimation.

Cumulative distributions of the sample means were generated to estimate the confidence interval for sample sizes from one to forty. Tables 10, 12 and 14 present the cumulative distributions for samples taken from the population with standard deviation of 0.5, 1.0 and 1.8 respectively. It should be noted that the values above and below the estimated (and actual) mean are not of equal value as are traditionally used for confidence intervals for means but are reciprocals. For example, the lower value of 0.8 is the reciprocal of the upper value of 1.25 and the lower value of 0.6 is the reciprocal of 1.7. One reason for this is that for relatively large confidence intervals, knowing the factor or multiple was thought to be more important than the equal value amount. Another reason is that even with high number of samples, the distribution of the means is positively skewed.

The cumulative distributions were used to produce probabilities/percentages of the 5,100 samples that were between the selected asymmetrical confidence intervals. Tables 11, 13 and 15 present the probabilities for the asymmetrical confidence intervals. It is notable that even with the population with the lowest standard deviation, more than ten samples are required to have a better than average chance of being within 10% of the population mean. Also, for the population with a standard deviation of 1.0 (the relative standard deviation closest to typical value for emission factors observed thus far), it requires more than 40 samples to produce an even chance that the value is within 10% of the true value. The number of emission tests to achieve these probabilities would imply that some people's expectation or belief that all "A" rated emission factors have this level of precision is unfounded. Another observation is with even with few unbiased samples, a precision of a factor of two (0.5 to 2) is reasonable. Therefore, except in situations where there is bias in the test method and/or the selection of the facilities tested, the assumption that an "E" rated emission factor is an order of magnitude estimate is also unfounded.

Providing users of emission factors a quantitative assessment of the accuracy of emission factors with "A," "B," "C," "D," or "E" ratings may make them comfortable about the use of the higher rated factors and skeptical when using a lower rated factor. However, any procedures developed would be very complex. Using the classical statistical approach to establish ratings using the *t* test of the mean would be an alternative. However, the confidence interval for each rating would have to be agreed upon. Adopting the more commonly used probabilities of 90%, 95% or 99% would relegate most factors to an

"E" rating unless very broad confidence intervals were specified. Using 50% probabilities would provide sufficient latitude to use the full range of existing letters. However, since the underestimation of the sample standard deviation differs from the normally distributed population, it is not clear that the classical methodology of establishing the confidence interval would be appropriate.

To assess the ability of the *t* test to estimate the confidence interval, the percentages generated in Tables 10, 12, and 14 were used to calculate a value based upon the average sample standard deviation and the number of samples used to establish the average mean. This calculated value was compared with the actual value used as the criteria. The areas in Tables 10, 12 and 14 with dark shading were within 10% in predicting the confidence interval. The areas with light shading were within 20% in predicting the confidence interval. Generally, for populations with standard deviations below 1.0, the 90% confidence intervals can be estimated using the *t* distribution and the sample standard deviation. These estimates will be within 10% for source test sample sizes of five or more. For populations with standard deviations of about 1.8 and higher, the use of the *t* distribution provides a poor estimator of the lower confidence interval. However, for sample sizes of five or more, the 80% upper confidence interval can be estimated within 10% and the 90% confidence interval can be estimated within 20%. For populations with standard deviations above 1.0, estimating confidence intervals using a Monte-Carlo simulation of a population with size is probably advisable and characteristics like those inferred by the available samples.

Knowledge about the distribution of sample means can provide valuable information about the uncertainty associated with estimating national emissions. This information can also be used to arrive at an estimate of the uncertainty for smaller areas that contain only part of the population described by the emission factor. Just as sample size and sample standard deviation can be used to estimate the confidence interval of the population mean, educated assumptions about the population distribution can provide information on the probabilities that the average of part of a population is within a given range. Assuming that the average emission factor from a sub-population of the facilities defined by an inventory area is independent of the estimated emission factor for the population of all facilities, the joint probability can be determined. The joint probability is determined by multiplying the two independent probabilities for a given confidence interval. For example, if 20 sources were used to develop an emission factor of 1.0 and a sample standard deviation of 0.9, then (as shown in Tables 6 and 7) it would be reasonable to assume that the population mean was 1.0 and the population standard deviation was 1.0. As a result, the probability of the factor being between 0.7 and 1.4 is 89% (Table 13). This would be the probability of estimating the entire populations emissions within this confidence interval. If on the other hand, the area being inventoried contains 10 sources that were not used to develop the emission factor, then the probability that these sources have an average emission factor between 0.7 and 1.4 is 74% (Table 12). The joint probability that the population emission factor is between 0.7 and 1.4 and that the average emission factor for the 10 sources is also between 0.7 and 1.4 is (84%)X(74%) = 62%. Note that although the probability of being within this range of confidence of the average emission factor for the 10 sources is 62%, the probability for this confidence interval for each of the 10 sources is only (89%)X(30%) = 22%.

These probabilities and confidence intervals can be used in assessing the reliability of an inventory, comparing the reliabilities of various portions of the inventory and evaluating follow up strategies for the least reliable parts of an emission inventory. This information would allow efforts to improve the inventory quality to focus on those sources with the greatest potential for improvement. The above information provides information for a simple example. Rather than using resources to improve the 62% reliability of emission estimates from ten sources, these resources would produce greater benefits if they were used to improve the emission estimates that are only 22% reliable.

CONCLUSIONS

This assessment of the uncertainties associated with the emission factors that are commonly used to assemble an inventory and the application of these emission factors to national or regional inventories suggest that there is a poor understanding of these uncertainties. Often, individuals assume that the emission factor provides a more precise estimator than is possible. In other cases, the developers of emission factors indicate that the factor is less precise than the data might show. This assessment does not address the potential for uncertainties due to biases in the data. However, this assessment does infer some ranges of uncertainty that may be inherent in many emission inventories that are routinely used to assess the control strategies in a given region. It would be beneficial if information about the number and variability of emission factors could be made available so that inventory preparers could assess quantitatively the uncertainties of the final emission inventory.

REFERENCES

1. *Compilation of Air Pollutant Emission Factors, Volume I: Stationary Point and Area Sources, Fifth Edition, AP-42*. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Research Triangle Park, NC, 1995.
2. *Procedures for Preparing Emission Factor Documents*, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Research Triangle Park, NC, 1997; EPA 454/R-95-015, pp 1-10.
3. *Quality Assurance Volume VI: Chapter 4, Evaluating the Uncertainty of Emission Estimates*, U.S. Environmental Protection Agency, Research Triangle Park, NC and Emission Inventory Improvement Program Quality Assurance Committee, July 1996.
4. Beck, L. L.; Peer R. L.; Bravo L. A.; Yan Y. 1994. "A Data Attribute Rating System." In *Proceedings of an International Specialty Conference, The Emission Inventory: Applications and Improvement*, Air & Waste Management Association: Pittsburgh, PA, 1994; pp 721-733.
5. Snedecor, G. W.; Cochran, W. G. *Statistical Methods, 8th Edition*, 1989, Iowa State University Press / Ames.
6. Bennett, K.; "Uncertainty Analysis of AP-42 Emission Factors." In *Proceedings of an International Specialty Conference, The Emission Inventory: Planning for the Future*, Air & Waste Management Association: Pittsburgh, PA, 1997, Poster Session.
7. *Emission Factor Documentation for AP-42 Section 2.4, Refuse Combustion*, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC, August 1993.
8. *Emission Factor Documentation for AP-42 Section 2.4, Municipal Solid Waste Landfills*, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Research Triangle Park, NC, August 1997.
9. *Emission Factor Documentation for AP-42 Section 11.6, Portland Cement Manufacturing*, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Research Triangle Park, North Carolina, May 1994.
10. *Emission Factor Documentation for AP-42 Section 11.17, Lime Manufacturing*, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Research Triangle Park, NC, April 1994.
11. *Emission Factor Documentation for AP-42 Section 1.3, Fuel Oil Combustion*, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Research Triangle Park, NC, January 1997.
12. Fitzimmons, T. 1997. U. S. Environmental Protection Agency, Research Triangle Park, NC, *personal communication*.

Table 1. Descriptive Statistics in normal space (and log space).

	Population Standard Deviations		
	0.5	1.0	1.8
Population Mean	1.0 (-0.104)	1.0 (-0.338)	1.0 (-0.81)
Standard Deviation	0.5 (0.473)	1.0 (0.834)	1.8 (1.28)
Skewness	1.5 (0.0)	3.4 (0.015)	6.4 (0.026)
Median	0.9 (-0.104)	0.7 (-0.339)	0.6 (-0.821)
Minimum Value	0.2 (-1.65)	0.05 (-2.91)	0.01 (-4.45)
Maximum Value	2.4 (1.48)	13.0 (2.57)	30.3 (3.41)

Table 2. Cumulative Distributions for Simulated Populations

	Population Standard Deviations		
	0.5	1.0	1.8
Value	Percentage of Population with Value Less than Stated		
0.2	0.05	0.5	27
0.33	1.8	20	40
0.5	10.6	33.6	53.9
0.6	19.5	41.9	59.5
0.7	29.7	49.2	64.0
0.8	40.1	55.6	67.7
0.9	49.9	61.0	70.9
1.0	58.7	65.8	73.6
1.1	66.3	69.8	76.0
2.25	77.5	75.0	78.9
1.4	82.3	79.1	81.4
1.7	90.9	85.1	85.2
2.0	95.4	89.2	87.9
3.0	99.4	95.8	93.1
5.0	100	99.0	97.0

Table 3. Underestimation of the Population Standard Deviation

	Number of Samples Drawn						
	2	5	10	15	20	30	40
Normal Distribution	33%	8%	4%	2%	1%	0.6%	0.4%
0.5	22%	9%	5%	3%	3%	2%	1%
1.0	38%	22%	15%	11%	9%	7%	5%
1.8	50%	36%	26%	21%	18%	15%	12%

Table 4. Descriptive Statistics for means of samples drawn from a population with mean of 1.0 and Standard Deviation of 0.5.

	Number of samples taken from population																		
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
Average	1.01	1	1	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
Std Dev	0.5	0.35	0.29	0.25	0.23	0.2	0.19	0.18	0.17	0.16	0.14	0.13	0.13	0.12	0.11	0.1	0.1	0.1	0.1
Median	0.9	0.94	0.96	0.98	0.98	0.99	0.99	0.99	0.99	1	1	1	1	1	1	1	1	1	1
Max	4.39	3.22	2.65	2.33	2.19	2.06	1.89	1.83	1.77	1.75	1.65	1.55	1.51	1.51	1.48	1.4	1.36	1.36	1.34
Min	0.19	0.25	0.38	0.44	0.44	0.49	0.48	0.53	0.56	0.56	0.58	0.61	0.64	0.67	0.66	0.69	0.75	0.77	0.79

Table 5. Descriptive Statistics for standard deviations of samples from a population with mean of 1.0 and Standard Deviation of 0.5

	Number of samples taken from population																	
	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
Average	0.373	0.416	0.439	0.453	0.461	0.465	0.468	0.472	0.474	0.479	0.482	0.484	0.485	0.487	0.489	0.491	0.493	0.49
Median	0.231	0.332	0.375	0.4	0.416	0.426	0.433	0.44	0.445	0.455	0.462	0.465	0.468	0.471	0.476	0.48	0.484	0.49
Max	2.355	2.158	1.914	1.671	1.541	1.429	1.391	1.31	1.25	1.171	1.103	1.15	1.098	1.046	0.944	0.906	0.902	0.88
Min	0	0	0.01	0.06	0.06	0.08	0.07	0.07	0.123	0.128	0.157	0.165	0.158	0.178	0.213	0.233	0.246	0.26

Table 6. Descriptive Statistics for means of samples drawn from a population with mean of 1.0 and Standard Deviation of 1.0

	Number of samples taken from population																		
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
Average	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
Std Dev	1	0.69	0.57	0.49	0.43	0.39	0.37	0.34	0.33	0.31	0.28	0.26	0.25	0.23	0.22	0.2	0.18	0.17	0.15
Skewness	3.44	2.3	1.82	1.61	1.4	1.23	1.14	1.11	1.08	1.02	0.9	0.84	0.84	0.79	0.78	0.66	0.61	0.57	0.53
Median	0.71	0.83	0.88	0.91	0.92	0.93	0.94	0.95	0.96	0.96	0.97	0.97	0.98	0.98	0.98	0.99	0.99	1	1
Max	13	7.48	5.23	4.49	3.76	3.2	3.06	3.11	3.01	2.95	2.51	2.59	2.32	2.15	2.3	2.04	1.85	1.74	1.72
Min	0.1	0.1	0.13	0.15	0.22	0.25	0.27	0.32	0.33	0.35	0.36	0.35	0.39	0.44	0.45	0.55	0.53	0.54	0.62

Table 7. Descriptive Statistics for standard deviations of samples from a population with mean of 1.0 and Standard Deviation of 1.0

	Number of samples taken from population																	
	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
Average	0.629	0.709	0.758	0.784	0.802	0.817	0.835	0.845	0.855	0.865	0.88	0.89	0.897	0.905	0.924	0.934	0.943	0.95
Median	0.33	0.494	0.579	0.628	0.662	0.688	0.715	0.732	0.747	0.769	0.793	0.81	0.823	0.836	0.863	0.88	0.895	0.91
Max	8.97	7.354	6.363	5.66	5.142	4.845	4.548	4.336	4.136	3.842	3.596	3.408	3.244	3.076	2.809	2.78	2.615	2.46
Min	0	0	0.03	0.05	0.05	0.09	0.132	0.132	0.129	0.171	0.197	0.224	0.237	0.236	0.238	0.29	0.303	0.29

Table 8. Descriptive Statistics for means of samples drawn from a population with mean of 1.0 and Standard Deviation of 1.8.

	Number of samples taken from population																		
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
Average	1	1.01	1.01	1.02	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1	1.01	1.01	1	1	1	1	1
Std Dev	1.83	1.28	1.05	0.92	0.8	0.74	0.69	0.65	0.61	0.58	0.53	0.49	0.47	0.44	0.41	0.37	0.33	0.31	0.29
Skewness	6.37	4.2	3.76	3.28	2.76	2.55	2.4	2.3	2.18	2.1	1.89	1.81	1.66	1.62	1.49	1.31	1.14	1.01	0.94
Median	0.44	0.62	0.71	0.76	0.8	0.82	0.84	0.86	0.87	0.88	0.9	0.91	0.92	0.93	0.93	0.94	0.95	0.96	0.96
Max	30.3	15.3	13.1	10.2	8.16	6.93	6.89	6.4	5.73	5.2	4.91	5	4.42	4.44	4.02	3.27	2.85	2.97	2.75
Min	0	0	0	0.1	0.1	0.11	0.1	0.14	0.15	0.15	0.17	0.2	0.23	0.22	0.21	0.33	0.36	0.33	0.33

Table 9. Descriptive Statistics for standard deviations of samples from a population with mean of 1.0 and Standard Deviation of 1.8

	Number of samples taken from population																	
	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
Average	0.92	1.045	1.138	1.18	1.232	1.277	1.31	1.335	1.356	1.4	1.428	1.464	1.488	1.502	1.532	1.56	1.582	1.61
Median	0.364	0.565	0.691	0.767	0.837	0.895	0.944	0.985	1.019	1.087	1.133	1.175	1.215	1.242	1.3	1.349	1.387	1.43
Max	21.41	17.26	15.02	13.45	12.22	11.34	10.91	10.4	9.946	9.173	8.518	8.033	8.681	8.249	7.424	6.802	6.309	6.11
Min	0	0	0.01	0	0.02	0.05	0.06	0.08	0.101	0.137	0.139	0.138	0.138	0.134	0.273	0.299	0.331	0.38

Table 10. Cumulative Distributions for means of samples drawn from a population with mean of 1.0 and Standard Deviation of 0.5.

Value	Number of samples taken from population.																		
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
0.2	0.1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
0.3333	1.8%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
0.5	10.6	3.2%	0.9%	0.4%	0.1%	0%%	0%%	0%%	0%%	0%%	0%%	0%	0%	0%	0%	0%	0%	0%	0%
0.6	19.5	9.0%	4.8%	2.2%	1.3%	0.5%	0.3%	0.1%	0.2%	0.1%	0.1%	0%	0%	0%	0%	0%	0%	0%	0%
0.7	29.7	19.3	12.5	8.5%	6.2%	4.0%	3.3%	2.2%	1.5%	1.1%	0.6%	0.4%	0.2%	0.2%	0.1%	0%	0%	0%	0%
0.8	40.1	31.1	25.1	21.0	17.1	13.7	12.3	10.5	9.3%	7.9%	5.7%	4.4%	3.4%	2.6%	1.9%	1.1%	0.6%	0.2%	0.1%
0.9	49.9	44.6	40.8	37.3	33.4	32.2	30.8	29.2	26.9	25.9	23.4	21.3	19.5	18.1	16.7	13.8	11.9	10.0	7.9%
1	58.7	56.7	55.7	54.7	53.7	53.6	52.6	52.2	51.5	51.7	50.8	50.8	50.2	50.0	50.4	48.8	49.0	48.9	48.4
1.1	66.3	67.1	68.3	69.2	70.1	70.7	71.5	72.0	73.0	74.3	75.2	76.3	77.9	79.3	79.8	83.1	84.6	86.1	87.4
1.25	77.5	79.1	82.5	84.2	86.2	87.9	90.0	91.1	91.9	92.6	94.0	95.3	96.1	97.3	97.5	98.7	99.2	99.6	99.8
1.4	82.3	87.5	91.3	92.8	94.4	95.7	96.8	97.6	98.2	98.3	99.0	99.4	99.6	99.8	100	100%	100%	100%	100%
1.7	90.9	96.0	97.7	98.6	99.3	99.5	99.7	99.9	99.9	99.9	100%	100%	100%	100%	100%	100%	100%	100%	100%
2	95.4	98.5	99.4	99.8	99.9	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
3	99.4	99.9	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
5	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table 11. Probability of a sample mean between stated values drawn from a population with mean 1.0 and Standard Deviation of 0.5.

Range	Number of samples taken from population.																		
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
.2 to 5	99.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
.333 to 3	98%	98%	99.4	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
0.5 to 2	85%	95%	99%	99.4	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
0.6 to 1.7	72%	87%	93%	96%	98%	99%	99.5	100	100	100	100	100	100	100	100	100	100	100	100
0.7 to 1.4	53%	68%	79%	84%	88%	92%	94%	95%	97%	97%	98%	99%	99.4	100	100	100	100	100	100
0.8 to 1.25	35%	48%	57%	63%	69%	74%	78%	81%	83%	85%	88%	91%	93%	95%	96%	98%	99%	99.4	100
0.9 to 1.1	16%	22%	27%	32%	37%	39%	41%	43%	46%	48%	52%	55%	58%	61%	63%	69%	73%	76%	79%

Table 12. Cumulative Distributions for means of samples drawn from a population with mean of 1.0 and Standard Deviation of 1.0.

	Number of samples taken from population.																		
Value	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
0.2	0.5%	1%	0.3%	0.1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
0.333	18%	7%	3%	2%	0.7%	0.3%	0.2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
0.5	34%	21%	14%	10%	7%	5%	3%	2%	2%	1%	0.6%	0.3%	0.3%	0.1%	0%	0%	0%	0%	0%
0.6	42%	30%	23%	19%	15%	12%	10%	8%	7%	6%	4%	3%	2%	1%	0.6%	0.3%	0.1%	0.0%	0%
0.7	49%	40%	34%	29%	26%	23%	20%	18%	16%	15%	12%	10%	8%	7%	6%	3%	2%	1%	0.9%
0.8	56%	48%	44%	39%	37%	35%	33%	30%	29%	27%	25%	23%	20%	18%	17%	13%	10%	8%	7%
0.9	61%	55%	52%	49%	48%	46%	46%	44%	42%	41%	40%	39%	37%	36%	34%	31%	29%	27%	25%
1	66%	62%	60%	58%	58%	57%	57%	57%	56%	56%	56%	55%	54%	54%	54%	53%	52%	51%	52%
1.1	70%	68%	66%	67%	66%	67%	67%	67%	68%	68%	69%	69%	69%	70%	71%	71%	72%	73%	74%
1.25	75%	75%	75%	76%	77%	78%	78%	80%	81%	81%	83%	84%	85%	86%	87%	89%	90%	92%	93%
1.4	79%	80%	81%	83%	84%	86%	87%	88%	89%	89%	91%	92%	93%	94%	95%	96%	97%	98%	99%
1.7	85%	88%	89%	92%	93%	95%	95%	96%	97%	97%	98%	98%	99%	99.2	99.4	99.6	99.8	100	100%
2	89%	92%	94%	96%	97%	98%	98%	99%	99%	99.3	99.5	99.7	99.8	99.9	99.9	100%	100%	100%	100%
3	96%	98%	99.1	99.5	99.6	99.9	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
5	99%	99.9	99.9	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table 13. Probability of a sample mean between stated values drawn from a population with mean 1.0 and Standard Deviation of 1.0.

	Number of samples drawn from population.																		
Range	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
.2 to 5	93%	99%	99.7	99.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
.333 to 3	78%	91%	96%	98%	99%	99.6	99.8	100	100	100	100	100	100	100	100	100	100	100	100
0.5 to 2	56%	72%	80%	86%	90%	93%	95%	96%	97%	98%	99%	99.4	99.5	99.8	99.9	100	100	100	100
0.6 to 1.7	43%	57%	66%	73%	79%	83%	85%	88%	90%	91%	94%	96%	97%	98%	99%	99.3	99.7	100	100
0.7 to 1.4	30%	40%	48%	53%	59%	63%	67%	70%	72%	74%	78%	83%	85%	87%	89%	93%	95%	97%	98%
0.8 to 1.25	19%	27%	31%	36%	40%	43%	46%	50%	52%	54%	57%	61%	65%	68%	70%	75%	80%	83%	86%
0.9 to 1.1	9%	13%	14%	17%	19%	20%	22%	23%	25%	27%	29%	30%	32%	35%	36%	40%	43%	46%	49%

Table 14. Cumulative Distribution for means of samples drawn from a population with mean of 1.0 and Standard Deviation of 1.8.

Value	Number of samples taken from population.																		
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
0.2	27%	13%	6%	3%	2%	1%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
0.333	42%	26%	18%	13%	10%	7%	6%	4%	3%	2%	1%	0.9%	0.5%	0.3%	0.3%	0%	0%	0%	0%
0.5	54%	40%	34%	29%	25%	21%	19%	17%	15%	13%	10%	8%	7%	5%	4%	3%	2%	1%	1%
0.6	60%	48%	43%	38%	34%	32%	29%	27%	24%	23%	19%	17%	15%	13%	12%	9%	7%	6%	4%
0.7	64%	55%	50%	46%	43%	41%	39%	37%	35%	34%	30%	28%	26%	25%	23%	20%	16%	14%	13%
0.8	68%	61%	57%	53%	51%	50%	48%	46%	45%	44%	41%	40%	39%	37%	35%	33%	30%	28%	26%
0.9	71%	66%	62%	60%	58%	57%	55%	55%	54%	53%	51%	50%	49%	48%	48%	45%	44%	43%	41%
1	74%	69%	67%	65%	64%	63%	63%	62%	61%	62%	60%	60%	59%	59%	59%	58%	58%	57%	57%
1.1	76%	73%	71%	70%	69%	69%	68%	68%	68%	68%	68%	68%	68%	68%	68%	69%	69%	69%	69%
1.25	79%	76%	76%	76%	75%	75%	75%	75%	76%	76%	77%	78%	78%	78%	78%	80%	81%	82%	82%
1.4	81%	80%	80%	80%	80%	81%	80%	81%	82%	83%	84%	84%	84%	85%	86%	88%	89%	89%	90%
1.7	85%	85%	86%	86%	87%	88%	88%	89%	90%	90%	91%	92%	92%	93%	93%	95%	96%	97%	98%
2	88%	88%	90%	90%	91%	92%	92%	93%	94%	94%	95%	96%	96%	96%	97%	98%	99%	99.3	100
3	93%	94%	96%	96%	97%	98%	98%	98%	98%	99%	99%	99.3	99.6	99.7	99.9	99.9	100%	100%	100
5	97%	98%	99%	99.2	99.5	99.6	99.7	99.9	99.9	100%	100%	100%	100%	100%	100%	100%	100%	100%	100

Table 15. Probability of a sample mean between stated values drawn from a population with mean 1.0 and Standard Deviation of 1.8.

Range	Number of samples drawn from population.																		
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40
0.2 to 5	70%	85%	92%	96%	97%	98%	99%	99.5	99.7	99.9	100	100	100	100	100	100	100	100	100
0.333 to 3	52%	69%	78%	84%	87%	91%	92%	94%	95%	96%	98%	98%	99.1	99.4	99.6	99.9	100	100	100
0.5 to 2	34%	40%	47%	52%	57%	60%	64%	66%	69%	71%	76%	78%	81%	83%	85%	89%	92%	94%	95%
0.6 to 1.7	26%	37%	43%	48%	53%	56%	59%	62%	65%	67%	72%	74%	77%	79%	82%	86%	89%	91%	93%
0.7 to 1.4	17%	25%	29%	35%	37%	39%	41%	44%	46%	49%	53%	56%	58%	61%	63%	68%	72%	75%	77%
0.8 to 1.25	11%	16%	19%	22%	24%	26%	27%	29%	31%	33%	36%	38%	39%	41%	43%	48%	50%	53%	56%
0.9 to 1.1	5%	7%	9%	10%	11%	12%	12%	13%	14%	15%	17%	18%	18%	20%	21%	23%	25%	25%	27%